神経回路網の基礎

甘利俊一 (理化学研究所 脳科学総合研究センター)

レクチャーノート作成　内田 肇（奈良先端科学技術大学院大学）

1　はじめに

生物学と数学の融合というすばらしい試みで大賛成だが，非常に難しいので，まだ10年や20年はかかるだろう．また，この試みが上手くいくか失敗するかはそのときになってわかるでしょう．物理学や工学の世界では数学的方法は成功したけれど，複雑な生物学の領域で数学的方法が本当に有効なのかはわからない．この数学的方法というのは，今ある物を意味しているわけでなく，人間の脳が持っているような数学的な論理的に物事を認識する方法のことです．このような数学的方法が解明されるのか，それとも，事実が複雑に雑多にあって個別に理解しておしまいとなる世界なのか．私は個人的には人間は物事を理解する仕方がちゃんとあると思っているので，この試みに賛同しています．

今回は，生物らしきものを題材にして，どういう数学ができるかという話をします．生物学というより数学オリエンテッドな話で，また，非常にクラシックな話です．

2　ニューロンのモデル

まず，ニューロンモデルを考える．ニューロンはイオンチャンネルのことなどを考え出すと非常に複雑だが，ここでは単純化し，外部から非常にたくさんの信号を受け取り，それらを総合して答えを出す素子だと考える．このとき，単純に入力を足し合わせるのではなく，それぞれの入力に重みをかけたものの和であるとする．重みというのは分子機構で決まるシナプスの結合効率であり，それは可塑的である．つまり学習によって重みは強くなったり弱くなったりなど，いろんな変化をする．

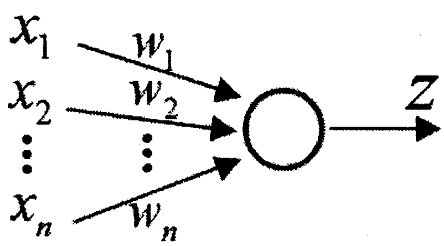

図 1

$$\mathbf{w}\cdot\mathbf{x} = w_1x_1 + \ldots + w_nx_n$$
$$Z == f(u) = f(\mathbf{w}\cdot\mathbf{x} - h)$$

入力を $\mathbf{x}$，重みを $\mathbf{w}$ としたとき，入力に重みをかけたものの和はベクトルの内積として $\mathbf{w}\cdot\mathbf{x}$ で表せる．このとき， $\mathbf{w}\cdot\mathbf{x}$ がある閾値 h を超えると興奮して電気パルスを発生する．ニューロンの電気パルスがあるかないかを 1 or 0 で表すとすると，正であれば 1，負であれば 0 を出す関数 $f(\cdot)$ を用いて表すことができる．これは，McCulloch-Pitts のニューロンモデルである．

入出力をパルスではなく 0 と 1 の間の実数値をとるようにしてやれば，いろいろな見方ができる．まず，1 つは，入出力の値をパルスが出る確率としてみるものである．この見方は，ニューロンは膜電位の熱雑音など，いろんな要因でゆらぎを伴っており確率的に動作をするからである．$f(\mathbf{w}\cdot\mathbf{x}-h)$ が高ければパルスを出す確率が高くなり，低ければパルスは殆ど出ない．

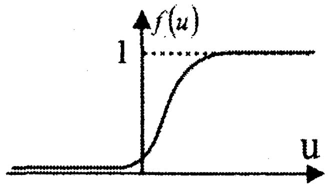

図 2

さらに見方を拡張し，同種のニューロンがたくさんあって同じような情報処理をやっており，それら全部が確率的に動作をすると考えると，$f(\mathbf{w}\cdot\mathbf{x}-h)$ の 1/2 のところは半分が興奮して半分が興奮していないことになる．つまり，ニューロン集団の発火頻度を表すとみることもできる．もう 1 つは，あるニューロンでパルスの発生が繰り返し起こっているときに，一定の時間スパンで平均してやるものであり，ニューロンの時間方向の発火頻度という見方である．

また，信号は連続時間で入ってくると考えることもでき，一番単純なのは

$$\tau\cdot\dot{u}=-u+\mathbf{w}\cdot\mathbf{x}-h$$
$$z=f(u)$$

というモデルである．$\mathbf{x}$ が時間とともにやってくるとき，その重み付け総和を表す u の変動でニューロンが発火するかどうかが決まる．u は膜電位のようなものだと考えてよい．単純に $\mathbf{w}\cdot\mathbf{x}$ の和だけでは増えるのみになってしまうので，減衰項 $-u$ も加えてある．入力 $\mathbf{x}$ が時間的に一定のとき，平衡状態を考えると $u=\mathbf{w}\cdot\mathbf{x}-h$ となり，離散時間モデルを連続時間へと拡張したモデルといえる．

3 1-layer network

次に相互に結合していないニューロンが並列に沢山あるときを考える．

$$\mathrm{sgn}\left(\sum_{j=1}^{n}w_{ij}x_j\right)$$
$$W=[\mathbf{w}_1\,\mathbf{w}_2\,...\,\mathbf{w}_n]^T$$
$$\mathbf{z}=T_w\mathbf{x}=\mathrm{sgn}(W\mathbf{x})$$

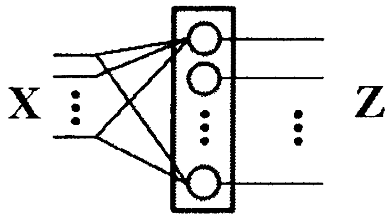

図 3

i 番目のニューロンは n 次元の入力 $\mathbf{x}$ を受けとり，重みベクトル $\mathbf{w}_i$ を使って出力 z_i を出力する．基本的には前章の 1 つのニューロンの場合と同じだが，式を簡単にするため出力を少し変え $\mathbf{w}_i\cdot\mathbf{x}$ が正のときは 1，負のときは -1 としている．また，$\mathbf{w}$ と $\mathbf{x}$ を調整することで省ける閾値 h もここでは考えない．ニューロンは n 個あるので n 次元の入力 $\mathbf{x}$ を受け n 次元の出力 $\mathbf{z}$ へと変換するモデルである．

この神経回路網は 2^n 個の信号を 2^n 個の信号のいずれかに移す変換であり，その法則は W によって決まる．変換の法則を決める W は無数にあるが，実際の神経回路網では，すべてが使われているわけではなく，そのなかの特殊なものが使われているかもしれない．では，どのような W を選んでやれば，どのような変換が可能となるのだろうか？

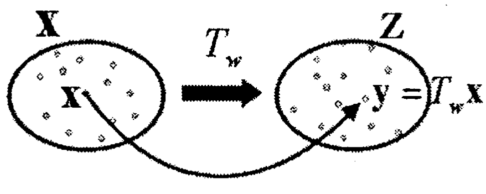

図 4

3.1　入力信号の差異の拡大（ランダム結合）

変換の法則を決める W がランダムであるときを考える．例えば，平均 0，分散 1 の正規分布から，それぞれ独立にとってくるとしたとき，取り得る W の空間には確率密度が入ってくる．この確率分布がある種の対称性をもっていれば，それに対応して $\mathbf{x}$ から $\mathbf{z}$ への変換に対称性が表れる．つまり，一見違う W を持っている神経回路網にも共通の法則が考えられる．今，W は独立同一分布からとってきているので，W の中の順番は問題にならなくなり，この強力な性質からいろんなことが伺える．例えば，入力 $\mathbf{x}$ がきたときに，それを平均したものを X とし出力も同様に平均し Z とする．

$$X = X(\mathbf{x}) = \frac{1}{n}\sum x_i$$
$$Z = Z(\mathbf{z}) = \frac{1}{n}\sum z_i$$

これらは，入力または出力の何割が発火しているかという発火率であると考えることができる．そうすると，入力の発火率 X が同じなら，ある出力の発火率 Z が出てくるというマクロな法則 $F(\cdot)$ が，すべての W について，ほぼ近似的に成立するだろうと想像できる．

$$Z = F(X) = F(X(\mathbf{x}))$$

Z は z_i の平均であるが，z_i は独立で同一分布な w_{ij} の関数なので，z_i も独立同一分布に従う確率変数である．独立同一分布に従う変数の非常に沢山の平均であれば対数の法則より期待値に近づく．つまり，ニューロンの数 n が十分に大きいとき，Z を求めるには z_i の期待値を計算すればよい．

$$Z = \mathrm{Prob}\{z_i = 1\} - \mathrm{Prob}\{z_i = -1\}$$
$$\mathrm{Prob}\{z_i = 1\} = \mathrm{Prob}\{\sum_{j=1}^{n} w_{ij}x_j > 0\}$$

期待値を求めるには $\mathrm{Prob}\{z_i = 1\}$ が計算できればよい．つまり，$\sum_{j=1}^{n} w_{ij}x_j$ が 0 以上になる確率を計算する．$\sum_{j=1}^{n} w_{ij}x_j$ は中心極限定理により $N(n\bar{w}X, n\sigma_w^2)$ に従うため，誤差関数を使って計算することができる．

$$Z = 2\mathrm{erf}\left(\frac{n\bar{w}X}{\sqrt{n\sigma_w^2}}\right) - 1$$

このとき出力の発火率 Z は入力の発火率 X の連続関数としてかけ，この関数は w_{ij} の平均 $\bar{w}$ や分散 σ_w^2 に依存して決まる．

単純に X が決まれば Z がわかるというだけでは面白くない．もし，X から少しずれた X' を入力として神経回路網に与えたとき，出力 Z' が Z と比較してどのように変化するかを考える．例えば X と X' が近いときは Z と Z' も近いのかを考えるということである．ここで，近さをはかるのにハミング距離を用いる．

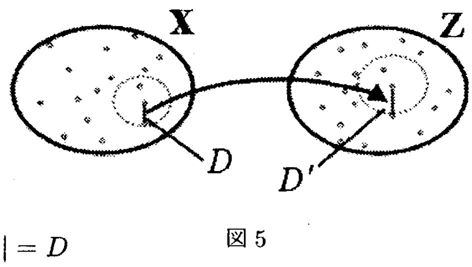

図 5

$$D(\mathbf{x}, \mathbf{x}') = \frac{1}{2n}\sum |x_i - x_i'| = D$$
$$D(\mathbf{z}, \mathbf{z}') = \frac{1}{2n}\sum |z_i - z_i'| = D'$$

このとき，Z と Z' の距離は $D' = D(\mathbf{z}, \mathbf{z}') = \mathrm{Prob}\{z_i \neq z'_i\}$ を計算してやればよく，

$$D' = \frac{2}{\pi} sin^{-1}(C\sqrt{D})$$

となる．これは，小さな D のときは大きな D' となるということであり，ランダムな神経回路網では，入力 X を少しずらすと出力 Z は大きくずれたものとなる．

これがどういう意味をもつかを考えると，信号の小さな違いを精密にしりたいときはランダムな結合にしてやればよいということになる．たとえば小脳であれば，苔状線維から顆粒細胞への結合というはランダムにみえる．ここで情報の差異を拡大し，それをプルキンエ細胞が学習し判別しているとみることができる．つまり学習をしやすくするために，このような仕掛けを作ったというのが小脳パーセプトロン仮説である．

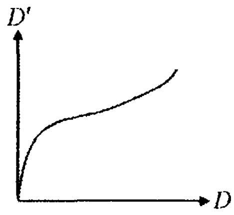

図 6

3.2 Associative memory

次に，W をランダムではなく意図的に決める．例えば，$\mathbf{s}_1$ という信号が来たときは $\mathbf{q}_1$ を出力し，$\mathbf{s}_2$ という信号が来たときは $\mathbf{q}_2$ を出力するというような連想記憶モデルのときである．ただし，統計的な性質をみたいので，入力 $\mathbf{s}$ と出力 $\mathbf{q}$ のセットはランダムに決める．このときの D と D' の関係をみると，入力が少し変わっても出力が殆ど変わらないものとなる．

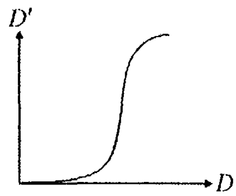

図 7

入力が少しぐらいでたらめでも正しい信号が再生できるという性質は連想記憶モデルに使える．海馬という記憶を整理し短期記憶をつかさどる領域では，このような仕掛けで情報を固定している可能性がある．例えば，MIT の利根川さんはマウスをノックアウトして海馬の特定の場所で特定のタンパクを作れなくし，上手い情報変換ができなくなっているというのをネズミの行動をみて示している．

4 Dynamics of Neuro-Ensembles

ニューロンはつながりにより情報のフィードバックがある．ニューロンの興奮は仲間のニューロンに伝わり，また仲間のニューロンに戻ってくる．このようなダイナミクスを考える．つまり，情報の変換というスタティックなモデルでなく，時間とともに変化するダイナミックなモデルで考える必要がある．

4.1 相互結合のある神経回路網の基本的性質

相互結合のある神経回路網の連続時間モデルを考える．

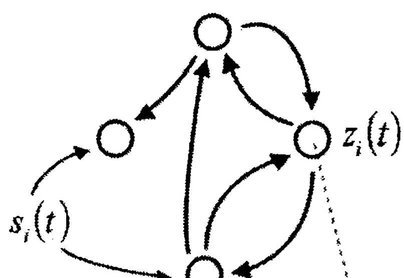

図 8

$$\tau \dot{u}_i(t) = -u_i + \sum_j w_{ij} z_j(t)$$
$$z_j(t) = f(u_j(t))$$
$$\tau \dot{\mathbf{u}} = W f(\mathbf{u}) + \mathbf{s} - \mathbf{u}$$

今までのモデルでは入力 $\mathbf{x}$ と出力 $\mathbf{z}$ は別だったが，ここでは各ニューロンの出力がそのまま次の入力となるモデルになっている．また，$\mathbf{s}$ は外部入力，$f(\cdot)$ はシグモイド状の関数である．このモデルを考えることで，ニューロン集団の何割が発火しているかという発火率が時間とともにどうやって発展していくかを考えることができる．

ここで，$\mathbf{u}$ の平均を U とし膜電位の平均だと考え，U の微分方程式を考える．

$$U = \frac{1}{n}\sum u_i$$
$$\dot{U} = Wf(U) + S - U$$

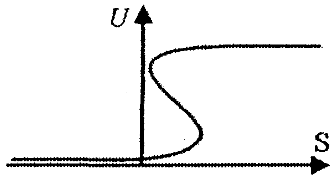

図 9

もし，W が小さければ $-U$ の項が勝ち，この微分方程式は単安定になる．逆に W が大きければ $Wf(U)$ の項が強くなり大きい値か小さい値に発散する方向に働くが，$f(\cdot)$ はシグモイド状であるので上限・下限ができ，双安定になる．この場合，外部入力 S によって U のとる状態が変わる．S が小さいときは U は小さい値で安定となり，S が大きいときは U は大きい値で安定となるが，S が中くらいの大きさのときは S の初期値に依存して U の値が決まり安定状態が決まり，ヒステリシスが生じる．このような性質はメモリー素子として使える．これを発展させ，一方は興奮性，一方は抑制性という 2 つの神経集団が相互に結合していると考たときは振動が起こることがわかっている．

4.2　相互結合がある場合の問題点

前節ではニューロン集団の出力がそのまま入力となるモデルにおいて，単純に膜電位の平均 U の微分方程式を考えたが，実は数学的には問題が生じている．その問題を議論するために，より簡単なランダム結合 W をもつ神経回路の離散時間モデルにもどる．離散時刻 t において，ニューロン集団 $\mathbf{x}$ の興奮・非興奮の状態を $\mathbf{x}_t$ で表すとすると，$\mathbf{x}_{t+1} = \mathrm{sgn}(W\mathbf{x}_t)$ という状態方程式で表せる．3 章ではニューロン集団の状態 $\mathbf{x}$ を平均したものを発火率 X として表したが，相互結合のあるモデルにおいても単純に発火率 X の状態方程式

$$X_{t+1} = F(X_t)$$

が成り立つのだろうか．実はここに問題がある．時刻 2 と 3 の X_t は以下のようになる．

$$X_2 = X(\mathbf{x}_2) = X(T_w\mathbf{x}_1)$$
$$X_3 = X(\mathbf{x}_3) = X(T_w\mathbf{x}_2) = X(T_wT_w\mathbf{x}_1)$$

$\mathbf{x}_1$ と T_w は無関係であるため中心極限定理が使え X_2 が求まる．しかし，X_3 を求めるときはどうだろうか．X_2 は T_w に依存して決まっているため，T_w と X_2 の間に関係が生まれ本当は中心極限定理が使えない．なにか別の方法を使って，T_w と X_2 の間に関係があるときでも，$F(\cdot)$ が成り立つことを証明しなければならない．

ここで弱い意味での成立と強い意味での成立を考える．真の発火率を $\widetilde{X}_t$，$F(\cdot)$ を用いて求めた発火率を X_t

$$\widetilde{X}_{t+1} = X(\mathbf{x}_t)$$
$$X_{t+1} = F(X_t)$$

とすると，

$$\text{強い意味での成立：} \lim_{n\to\infty} \sup_t E[\|X_t - \widetilde{X}_t\|^2] = 0$$

$$\text{弱い意味での成立：} \lim_{n\to\infty} E[\|X_t - \widetilde{X}_t\|^2] = 0$$

を考える．つまり，時刻 t を固定して $n \to \infty$ としたときに成立することを証明するのが弱い意味である．W が正規分布ではなく安定分布に従うとき，弱い意味での証明はされ，また非常に単純な確率分布に従う W を考えたときは強い意味でも証明されている．結局，この問題は未だに限られた状況でしか証明されていない．

参考文献

[1] 甘利俊一, 神経回路網の数理, 産業図書 (1978)